

Time-dependent behaviour of an alternating service queue

Maria Vlasiou and Bert Zwart

H. Milton Stewart School of Industrial & Systems Engineering
Georgia Institute of Technology
765 Ferst Drive, NW
Atlanta, Georgia 30332-0205

November 24, 2006

vlasiou@gatech.edu, bertzwart@gatech.edu

Abstract

We consider a model describing the waiting time of a server alternating between two service points. This model is described by a Lindley-type equation. We are interested in the time-dependent behaviour of this system and derive explicit expressions for its time-dependent waiting-time distribution, the correlation between waiting times, and the distribution of the cycle length. Since our model is closely related to Lindley's recursion, we compare our results to those derived for Lindley's recursion.

1 Introduction

Consider a system consisting of one server and two service points. At each service point there is an infinite queue of customers that needs to be served. The server alternates between the service points, serving one customer at a time. Before being served by the server, a customer must undergo first a preparation phase, which starts immediately after the server has completed service at that particular service point and has moved to the next one. Thus the server, after having finished serving a customer at one service point, may have to wait for the preparation phase of the customer at the other service point to be completed. Let W_n be the time the server has to wait before he can start serving the n -th customer. If B_n is the preparation time of the n -th customer and A_n is the service time of the n -th customer, then it is easy to see that W_n can be defined recursively by

$$W_{n+1} = \max\{0, B_{n+1} - A_n - W_n\}, \quad n \geq 1. \quad (1.1)$$

We assume that $\{A_n\}$ and $\{B_n\}$ are two sequences of i.i.d. random variables that are also mutually independent. A natural assumption in the above setting is that $W_1 = B_1$. This alternating service model occurs in many applications. For example, this strategy is followed by surgeons performing eye surgeries. Another example where (1.1) occurs comes from inventory theory, in particular from the analysis of two-carousel systems. This application is considered in [9, 10, 13]. Specifically, in Park *et al.* [13], the goal is to derive the steady-state waiting time distribution under specific assumptions on the distributions of A_n and B_n that are relevant to the carousel application considered. These results are extended in [14, 15, 16, 17], where the main focus is on

the steady-state distribution of the waiting time. Contrary to the above-mentioned work, this paper focuses on the *time-dependent* behaviour of the process $\{W_n\}$. We are particularly interested in the distribution of W_n for any n , in the covariance between W_n and W_{n+k} , and in the distribution of the length of $C = \inf\{n \geq 1 : W_{n+1} = 0 \mid W_1 = 0\}$.

It should be clear to the reader that the recursion (1.1) is, up to the minus sign in front of W_n , equal to Lindley's recursion, which is one of the most important and well-studied recursions in applied probability. In Lindley's recursion, A_i represents the interarrival time between customers i and $i + 1$ and B_i the service time of customer i ; see for example [2] and [7] for a comprehensive description. Apart from trying to compute the above-mentioned quantities, another goal of our research is to compare the complexity between the two models when it comes to the analysis of the time-dependent behaviour. It is well known that for Lindley's recursion, the time-dependent waiting-time distribution is determined by the solution of a Wiener-Hopf problem, see for example [2] and [7]. In Section 2.1, we explore the possibility of an analytic approach. We derive an integral equation for the generating function of the distribution of W_n , and conclude that this integral equation is equivalent to a *generalised* Wiener-Hopf equation, which cannot be solved in general. This makes it appear that (1.1) may have a more complicated time-dependent behaviour than Lindley's recursion. However, a point we make in this paper is that this is not necessarily the case. Using probabilistic arguments, we analyse the distribution for exponential and phase-type preparation times. The expressions we obtain are remarkably explicit. Thus, Equation (1.1) is a rare example of a stochastic model which allows for an explicit time-dependent analysis. The reason is that, if B_1 has a mixed-Erlang distribution, we can completely describe (1.1) in terms of the evolution of a finite-state Markov chain; for details we refer to Section 5. We illustrate the difference in complexity between (1.1) and Lindley's recursion in Section 4.

We obtain similar explicit results for the distribution of the cycle length C . In particular, we do not need to resort to the usage of generating functions, as is necessary when analysing the corresponding quantity in Lindley's recursion. Note that the interpretation of C for our model is completely different from the corresponding quantity for Lindley's recursion. There, C represents the number of customers that arrived during a busy period. In our setting, C represents the number of pauses a server has until he needs to serve two consecutive customers without any pause. In this sense C can be seen as a “non-busy period”.

If one cannot determine the limiting distribution function of the waiting time analytically, but wants to obtain an estimation from a simulation of the system, then two relevant questions are how big is the variance of the mean of a sample of successive waiting times, and how long the simulation should run. For the determination of the magnitude of this variance it is necessary and sufficient to know the covariance function of the process $\{W_n\}$. In Section 5.3 we compute the covariance function for phase-type preparation time distributions, while we review existing results for the covariance function for Lindley's equation in Section 4.3.

This paper is organised as follows. In Section 2, we investigate several qualitative properties of $\{W_n\}$. Specifically, we show that, in general, the generating function of the distribution of W_n is determined by a generalised Wiener-Hopf equation and this equation is a contraction mapping. In addition, we investigate several properties of the covariance function. We show that, under weak assumptions, this function is of an alternating sign. Furthermore, we obtain an upper bound for its absolute value, from which we conclude that the covariance function converges to zero geometrically fast. Section 3 presents a detailed treatment of the special case of exponentially distributed preparation times, which can be seen as an analogue of the G/M/1 queue. The analysis

in this case is particularly tractable and leads to explicit and fairly simple expressions for the distribution of W_n , the distribution of C , and the covariance between W_n and W_{n+k} . These expressions are compared to the corresponding quantities for the G/M/1 queue in Section 4. This section also gives a representation for the transient G/M/1 waiting time distribution which, to the best of our knowledge, is a new result. In Section 5 we return to the alternating service model and extend the results of Section 3 to Erlang and mixed-Erlang preparation times.

2 Some qualitative properties

The goal of this section is to derive a number of qualitative results for the process $\{W_n\}$. In particular, in Section 2.1, we derive an equation that determines the distribution of W_n , and study some of its features, while in Section 2.2, we derive various properties of the covariance between W_1 and W_{1+k} .

As mentioned before, we are interested in the stochastic recursion

$$W_{n+1} = \max\{0, X_{n+1} - W_n\}, \quad n \geq 1, \quad (2.1)$$

with $X_n = B_{n+1} - A_n$. Let X be an i.i.d. copy of X_1 and let F_X be the distribution function of X . We assume that F_X is continuous and that $F_X(0) = \mathbb{P}[X \leq 0] \in (0, 1)$. Under these assumptions, the process $\{W_n\}$ is regenerative, with the epochs n where $W_n = 0$ being the regeneration points. Furthermore, it is shown in [14] that W_n converges in distribution to a random variable W , which is the limiting waiting time of the server. Let C_1 be first time after time $n = 1$ that a zero waiting time occurs, i.e., $C_1 = \inf\{n \geq 1 : W_{n+1} = 0\}$. Define also the generic regeneration cycle as $C = \inf\{n \geq 1 : W_{n+1} = 0 \mid W_1 = 0\}$. Note that, in general, C and C_1 have different distributions. Moreover, for a random variable Y we denote its distribution (density) by F_Y (f_Y). We can now proceed with the analysis.

2.1 An integral equation

A classical approach to Lindley's recursion when studying the time-dependent distribution of the waiting time is to consider the generating function of the Laplace transform of W_n ; see, for example, Cohen [7]. In this light, we first consider the generating function of the waiting times.

Therefore, define for $|r| < 1$ and $x \geq 0$ the function

$$H(r, x) = \sum_{n=0}^{\infty} r^n \mathbb{P}[W_{n+1} \leq x].$$

Notice that, since the distribution function of X_{n+1} is continuous, (2.1) yields

$$\mathbb{P}[W_{n+1} \leq x] = 1 - \mathbb{P}[X_{n+1} - W_n \geq x] = 1 - \int_x^{\infty} \mathbb{P}[W_n \leq y - x] dF_X(y).$$

Consequently, for the generating function we have that

$$\begin{aligned}
H(r, x) &= \mathbb{P}[W_1 \leq x] + \sum_{n=1}^{\infty} r^n \mathbb{P}[W_{n+1} \leq x] \\
&= \mathbb{P}[W_1 \leq x] + \frac{r}{1-r} - \sum_{n=1}^{\infty} r^n \int_x^{\infty} \mathbb{P}[W_n \leq y-x] dF_X(y) \\
&= \mathbb{P}[W_1 \leq x] + \frac{r}{1-r} - r \int_x^{\infty} H(r, y-x) dF_X(y).
\end{aligned} \tag{2.2}$$

Note that this equation is similar to the equation derived in [14, Section 4] for the limiting waiting time distribution W . It would be interesting of course to be able to solve equation (2.2) in general. To see to which extent this is possible, we investigate various properties of this equation.

We first show that (2.2) can be reduced to a generalised Wiener-Hopf equation, assuming that W_1 and X have densities f_{W_1} and f_X on $(0, \infty)$. Under these assumptions, we see from (2.2) that $H(r, x)$ has a derivative $h(r, x)$ on $(0, \infty)$; therefore, by differentiating with respect to x , (2.2) yields

$$\begin{aligned}
h(r, x) &= f_{W_1}(x) + rH(r, 0)f_X(x) + r \int_x^{\infty} h(r, y-x)f_X(y)dy \\
&= f_{W_1}(x) + rH(r, 0)f_X(x) + r \int_0^{\infty} h(r, u)f_X(u+x)du.
\end{aligned} \tag{2.3}$$

Notice that $h(r, x) = \int_0^{\infty} h(r, y)\delta(x-y)dy$, with δ being the Dirac δ -function. Combining this with (2.3), we obtain that

$$\int_0^{\infty} h(r, y)[\delta(x-y) - rf_X(x+y)]dy = rH(r, 0)f_X(x) + f_{W_1}(x).$$

This equation is equivalent to a generalised Wiener-Hopf equation; see Noble [11, p. 233]. It is shown there that such equations can sometimes be solved, but a general solution, as is possible for the classical Wiener-Hopf problem (arising in Lindley's recursion), seems to be absent.

The fact that we are dealing with a generalised Wiener-Hopf equation could indicate that deriving the distribution of W_n for our model may be more complicated than for Lindley's recursion. One point we make in this paper is that this is not necessarily the case.

The integral equation (2.2) has the following property, which is proven to be valuable in overcoming the difficulties arising by the fact that we are dealing with a generalised Wiener-Hopf equation. We shall show that the function H is the fixed point of a contraction mapping. To this end, consider the space $\mathcal{L}^{\infty}([0, \infty))$, i.e., the space of measurable and bounded functions on the real line with the norm

$$\|F\| = \sup_{x \geq 0} |F(x)|.$$

In this space we define the mapping \mathcal{T}_r by

$$(\mathcal{T}_r F)(x) = \mathbb{P}[W_1 \leq x] + \frac{r}{1-r} - r \int_x^{\infty} F(y-x) dF_X(y).$$

Then, for two arbitrary functions F_1 and F_2 in this space we have

$$\begin{aligned}
\|(\mathcal{T}_r F_1) - (\mathcal{T}_r F_2)\| &= \sup_{x \geq 0} |(\mathcal{T}_r F_1)(x) - (\mathcal{T}_r F_2)(x)| \\
&= \sup_{x \geq 0} \left| r \int_x^\infty [F_2(y-x) - F_1(y-x)] dF_X(y) \right| \\
&\leq |r| \sup_{x \geq 0} \int_x^\infty \sup_{t \geq 0} |F_2(t) - F_1(t)| dF_X(y) \\
&= |r| \|F_1 - F_2\| \sup_{x \geq 0} (1 - F_X(x)) \\
&= |r| \mathbb{P}[X > 0] \|F_1 - F_2\|.
\end{aligned}$$

Since $|r| < 1$ and $\mathbb{P}[X > 0] < 1$, we see that \mathcal{T}_r is a contraction mapping on $\mathcal{L}^\infty([0, \infty))$ with contraction coefficient $|r|\mathbb{P}[X > 0]$. Thus, iterating \mathcal{T}_r leads to a numerical approximation of H .

Summarising the above, we see that we can either find H exactly by solving a generalised Wiener-Hopf equation, or in a computational manner by iterating the mapping \mathcal{T}_r . If this first step is successful, then one may invert H exactly or with a computational method (by applying the Fast Fourier Transform) to obtain values for $\mathbb{P}[W_n \leq x]$. Generalised Wiener-Hopf equations are not solvable in general, and in the next sections we give a direct approach which leads to explicit solutions for $\mathbb{P}[W_n \leq x]$, rather than its generating function H . However, we first investigate the covariance function.

2.2 The covariance function

In this section, we only assume that $\mathbb{P}[X > 0] \in (0, 1)$; the distribution of X need not be continuous unless stated otherwise. Let $c(k) = \text{cov}[W_1, W_{1+k}]$ be the covariance between the first and the $(k+1)$ -st waiting time. In Theorem 1 we show that the covariance between two waiting times is of alternating sign, while in Theorem 2 we bound the absolute value of $c(k)$ to conclude that the covariance function converges to zero geometrically fast. The result of Theorem 1 is an expected effect of the non-standard sign of W_n in Recursion (1.1), while the conclusion we draw from Theorem 2 reinforces the results of Vlasiou [14], where it is shown that $\mathbb{P}[W_n \leq x]$ converges geometrically fast to F_W . We proceed with stating Theorem 1.

Theorem 1. *The covariance function $c(k)$ is non-negative if k is even and non-positive if k is odd. If in addition, X has a strictly positive density on an interval (a, b) , $0 < a < b$, and $W \stackrel{\mathcal{D}}{=} W_1$, then $c(k) > 0$ if k is even, and $c(k) < 0$ if k is odd.*

In order to prove this theorem we shall first prove the following lemma, which is a variation of a result by Angus [1].

Lemma 1. *Let Y be a random variable and f a non-decreasing (non-increasing) function defined on the range of Y . Then, provided the expectations exist,*

$$\text{cov}[Y, f(Y)] \geq 0 \quad (\text{cov}[Y, f(Y)] \leq 0).$$

Furthermore, if Y is not deterministic and

$$\mathbb{P}[Y \in \{y : f(y) \text{ strictly increasing (decreasing) in } y\}] > 0,$$

then

$$\text{cov}[Y, f(Y)] > 0 \quad (\text{cov}[Y, f(Y)] < 0).$$

Proof. We prove this lemma only for f being non-decreasing. The proof for non-increasing f follows analogously. We use the same argument as Angus [1]. Let Z be an i.i.d. copy of Y . So, if f is non-decreasing, then we have that

$$(Y - Z)(f(Y) - f(Z)) \geq 0.$$

Furthermore, let I_Y be the subset of the domain of f where the function is strictly increasing, i.e., $I_Y = \{y : f(y) \text{ strictly increasing in } y\}$. Then, if $\mathbb{P}[Y \in I_Y] > 0$, we have that

$$\mathbb{P}[(Y - Z)(f(Y) - f(Z)) > 0] > 0.$$

By taking expectations, and using the fact that Y and Z are i.i.d. we obtain

$$\mathbb{E}[(Y - Z)(f(Y) - f(Z))] = 2 \text{cov}[Y, f(Y)],$$

which is non-negative, and strictly positive if $\mathbb{P}[Y \in I_Y] > 0$. □

Proof of Theorem 1. For $k = 0$ the statement is trivial. For any fixed integer $k > 0$ we condition on the event that $X_i = x_i \in \mathbb{R}$, for $i = 2, \dots, k + 1$. Conditionally upon this event, we recursively define the functions g_i as follows;

$$g_1(w) = w \quad \text{and} \quad g_{i+1}(w) = \max\{0, x_{i+1} - g_i(w)\}, \quad i = 1, \dots, k.$$

It can easily be shown now that g_1 is non-decreasing, g_2 is non-increasing and, by iterating, that g_i is non-increasing if i is even, and non-decreasing if i is odd.

Let the first waiting time be fixed; that is, $W_1 = w_1$. Then, it is clear that, for all $i = 1, \dots, k + 1$, the i -th waiting time will be equal to $g_i(w_1)$, cf. Recursion (1.1). Now, write

$$\text{cov}[W_1, W_{1+k}] = \int \cdots \int_{x_2 \in \mathbb{R}, \dots, x_{k+1} \in \mathbb{R}} \text{cov}[W_1, g_{k+1}(W_1)] d\mathbb{P}[X_2 \leq x_2, \dots, X_{k+1} \leq x_{k+1}].$$

From Lemma 1, we obtain that $\text{cov}[W_1, g_{k+1}(W_1)] \geq 0$ if k is even and that $\text{cov}[W_1, g_{k+1}(W_1)] \leq 0$ if k is odd. This concludes the first part of the theorem.

Assume now that X has a strictly positive density on (a, b) ; therefore, for all $a_1, a_2 \in (a, b)$, with $0 < a_1 < a_2$, we have that $\mathbb{P}[X \in (a_1, a_2)] > 0$. We know already that for any set of fixed constants $\{x_i\}, i = 2, \dots, k + 1$, the functions g_i are monotone (i.e., either non-decreasing or non-increasing). Moreover, observe that if these constants have the property that $x_{k+1} > x_k > \cdots > x_2$, then g_{k+1} is strictly monotone in $(0, x_2)$.

Furthermore, since $\mathbb{P}[X \in (a_1, a_2)] > 0$ and $W \stackrel{\mathcal{D}}{=} W_1$, we have that

$$\mathbb{P}[W \in (a_1, a_2)] = \mathbb{P}[\max\{0, X - W\} \in (a_1, a_2)] \geq \mathbb{P}[W = 0] \mathbb{P}[X \in (a_1, a_2)] > 0,$$

which implies that $\mathbb{P}[W_1 \in (a_1, a_2)] > 0$ for all $a_1, a_2 \in (a, b)$, with $0 < a_1 < a_2$. So we have that if $x_2 > a$, then

$$\mathbb{P}[W \in (0, x_2)] \geq \mathbb{P}[W_1 \in (a, x_2)] > 0,$$

which can be rewritten as

$$\mathbb{P}[W \in \{w : g_{k+1}(w) \text{ strictly monotone in } w\}] > 0.$$

Therefore, by Lemma 1 we have that $\text{cov}[W_1, g_{k+1}(W_1)] > 0$ (< 0) if k is even (odd).

Now, let S be the subset of \mathbb{R}^k defined as follows,

$$S = \{(x_2, x_3, \dots, x_{k+1}) : x_{k+1} > x_k > \dots > x_2 > a\},$$

and let S^c be its complement. Then

$$\begin{aligned} \text{cov}[W_1, W_{1+k}] = & \int \cdots \int_{(x_2, x_3, \dots, x_{k+1}) \in S} \text{cov}[W_1, g_{k+1}(W_1)] d\mathbb{P}[X_2 \leq x_2, \dots, X_{k+1} \leq x_{k+1}] + \\ & \int \cdots \int_{(x_2, x_3, \dots, x_{k+1}) \in S^c} \text{cov}[W_1, g_{k+1}(W_1)] d\mathbb{P}[X_2 \leq x_2, \dots, X_{k+1} \leq x_{k+1}]. \end{aligned} \quad (2.4)$$

We know that the second integral at the right-hand side of (2.4) is greater than or equal to zero if k is even and less than or equal to zero if k is odd. It remains to show that the first integral at the right-hand side of (2.4) is strictly positive if k is even and strictly negative if k is odd. Since we integrate over the set S , we have shown that $\text{cov}[W_1, g_{k+1}(W_1)] > 0$ (< 0) if k is even (odd). So it suffices to show that $\mathbb{P}[S'] > 0$, where

$$S' = \{(X_2, X_3, \dots, X_{k+1}) \in S\} = \{X_{k+1} > X_k > \dots > X_2 > a\}.$$

Indeed, take a partition $\{a_i\}$ of (a, b) such that $a_i = a + [i(b-a)]/k$, $i = 0, \dots, k$. Then we have that

$$\begin{aligned} \mathbb{P}[X_{k+1} > X_k > \dots > X_2 > a] & \geq \\ \mathbb{P}[X_{k+1} \in (a_{k-1}, b); X_k \in (a_{k-2}, a_{k-1}); \dots; X_2 \in (a, a_1)] & = \\ \prod_{i=2}^{k+1} \mathbb{P}[X_i \in (a_{i-2}, a_{i-1})] & > 0, \end{aligned}$$

since X has a strictly positive density on (a, b) . □

This technique can be also applied to other stochastic recursions; for example, for Lindley's recursion the above argument shows under weak assumptions that the covariance between the waiting time of customer 1 and $k+1$ is strictly positive. Having seen that the correlations have alternating sign, we now turn to the question of the behaviour of the covariance function $c(k)$ for large k .

Theorem 2. *For every value of k we have that*

$$|c(k)| \leq \mathbb{E}[W_1] \mathbb{E}[X | X > 0] \mathbb{P}[X > 0]^k.$$

We see that $c(k)$ converges to zero geometrically fast in k . This is consistent with the fact that the distribution of W_n converges geometrically fast to that of W , cf. Vlasiou [14].

Proof. As in [14], we use a coupling argument. Define $T = \inf\{j \geq 1 : X_{j+1} \leq 0\}$. We write $W_{k+1} = W_{k+1}(W_1)$ to stress the fact that W_{k+1} is a function of W_1 . We see that

$$\begin{aligned} c(k) &= \int_0^\infty w (\mathbb{E}[W_{k+1}(w)] - \mathbb{E}[W_{k+1}(W_1)]) d\mathbb{P}[W_1 \leq w] \\ &= \int_0^\infty w (\mathbb{E}[W_{k+1}(w) - W_{k+1}(W_1)]) d\mathbb{P}[W_1 \leq w] \\ &= \int_0^\infty w (\mathbb{E}[(W_{k+1}(w) - W_{k+1}(W_1)) ; T > k]) d\mathbb{P}[W_1 \leq w]. \end{aligned}$$

The last equality holds since $W_{k+1}(w) = W_{k+1}(W_1)$ if $T \leq k$, i.e. T is a coupling time. Now, note that $|W_{k+1}(w) - W_{k+1}(W_1)| \leq X_{k+1}^+$ and that $T > k$ if and only if $X_2 > 0, \dots, X_{k+1} > 0$, so that

$$\mathbb{E}[|W_{k+1}(w) - W_{k+1}(W_1)| ; T \geq k] \leq E[X_{k+1} \mid X_{k+1} > 0] \mathbb{P}[X_1 > 0]^k.$$

We conclude that

$$|c(k)| \leq \int_0^\infty w E[X_{k+1} \mid X_{k+1} > 0] \mathbb{P}[X_1 > 0]^k d\mathbb{P}[W_1 \leq w].$$

Evaluating this integral yields the assertion. \square

In Section 4 we shall compare these results on the covariance function with known results for Lindley's recursion.

3 Exact solution for exponential preparation times

In this section we analyse the alternating service queue under the assumption that the preparation times B_i , $i \geq 1$, have an exponential distribution with rate μ . In the first part, we derive an explicit expression for the distribution of W_n . Later on, we derive the distribution of the cycle length C , and in Section 3.3 we compute the covariance between W_n and W_{n+k} . We define the Laplace-Stieltjes transform of A_1 by $\alpha(\cdot)$.

3.1 The time-dependent distribution

Although in the alternating service example, the choice $W_1 = B_1$ is natural, we would like to allow any initial condition. Therefore, we assume that $W_1 = w_1$ unless stated otherwise. Throughout this section, all probabilities are conditioned on this event. We first analyse the distribution of W_2 . Write for $x \geq 0$,

$$\mathbb{P}[W_2 > x] = \mathbb{P}[B_2 > A_1 + w_1 + x] = \int_0^\infty e^{-\mu(y+w_1+x)} dF_A(y) = e^{-\mu(x+w_1)} \alpha(\mu). \quad (3.1)$$

We see that $\mathbb{P}[W_2 > x \mid W_2 > 0] = e^{-\mu x}$, that is, the distribution of W_2 is a mixture of a mass at zero and the exponential distribution with rate μ . In order to compute the distribution of W_{n+1} , $n \geq 2$, observe that

$$\mathbb{P}[W_{n+1} > x] = \mathbb{P}[W_{n+1} > x \mid W_n = 0] \mathbb{P}[W_n = 0] + \mathbb{P}[W_{n+1} > x \mid W_n > 0] \mathbb{P}[W_n > 0]. \quad (3.2)$$

To calculate all terms that appear in (3.2) we need to compute the distribution of W_{n+1} conditioned on the length of the previous waiting time. To this end, we have that for $n \geq 2$,

$$\begin{aligned}\mathbb{P}[W_{n+1} > x \mid W_n = w] &= \mathbb{P}[B_{n+1} - A_n - w > x \mid W_n = w] \\ &= \int_0^\infty \mathbb{P}[B_{n+1} > x + y + w \mid W_n = w] dF_A(y) \\ &= \int_0^\infty e^{-\mu(x+w)} e^{-\mu y} dF_A(y) = e^{-\mu(x+w)} \alpha(\mu).\end{aligned}\tag{3.3}$$

For $w = 0$ we readily have the first term at the right-hand side of (3.2), i.e.,

$$\mathbb{P}[W_{n+1} > x \mid W_n = 0] = e^{-\mu x} \alpha(\mu), \quad n \geq 2.\tag{3.4}$$

Another implication of (3.3) is that for $n \geq 2$,

$$\mathbb{P}[W_{n+1} > x \mid W_{n+1} > 0, W_n = w] = \frac{\mathbb{P}[W_{n+1} > x \mid W_n = w]}{\mathbb{P}[W_{n+1} > 0 \mid W_n = w]} = e^{-\mu x}.\tag{3.5}$$

A straightforward conclusion is that

$$\mathbb{P}[W_{n+1} > x \mid W_{n+1} > 0] = e^{-\mu x}.\tag{3.6}$$

Thus, the distribution of W_{n+1} , provided that W_{n+1} is strictly positive, is exponential and independent of the length of the previous waiting time.

We can extend (3.5) to the following more general property. For any event E of the form $E = \{W_2 \in S_2, \dots, W_n \in S_n\}$, with measurable $S_k \subseteq [0, \infty)$, $2 \leq k \leq n$, we have that

$$\mathbb{P}[W_{n+1} > x \mid E, W_{n+1} > 0] = e^{-\mu x}.\tag{3.7}$$

To see this, write

$$\begin{aligned}\mathbb{P}[W_{n+1} > x \mid E, W_{n+1} > 0] &= \frac{\mathbb{P}[W_{n+1} > x; E, W_{n+1} > 0]}{\mathbb{P}[E, W_{n+1} > 0]} \\ &= \frac{\mathbb{P}[W_{n+1} > x; E]}{\mathbb{P}[W_{n+1} > 0; E]} = \frac{\mathbb{P}[W_{n+1} > x \mid E]}{\mathbb{P}[W_{n+1} > 0 \mid E]}.\end{aligned}$$

Furthermore, for $x \geq 0$,

$$\begin{aligned}\mathbb{P}[W_{n+1} > x \mid E] &= \int_0^\infty \mathbb{P}[W_{n+1} > x \mid W_n = w] d\mathbb{P}[W_n \leq w \mid E] \\ &= \int_0^\infty \int_0^\infty e^{-\mu(x+w+y)} dF_A(y) d\mathbb{P}[W_n \leq w \mid E] \\ &= e^{-\mu x} \alpha(\mu) \mathbb{E}[e^{-\mu W_n} \mid E],\end{aligned}$$

which directly proves (3.7), if we divide by the equation resulting for $x = 0$. Thus, the distribution of W_n is a mixture of a mass at zero and the exponential distribution with rate μ . This property is valid for all $n \geq 2$; for $n = 2$ it was shown below (3.1).

We now get back to Equation (3.2). Write, for $n \geq 2$,

$$\begin{aligned}\mathbb{P}[W_{n+1} > x \mid W_n > 0] &= \mathbb{P}[W_{n+1} > x, W_{n+1} > 0 \mid W_n > 0] \\ &= \mathbb{P}[W_{n+1} > x \mid W_{n+1} > 0, W_n > 0] \mathbb{P}[W_{n+1} > 0 \mid W_n > 0] \\ &= e^{-\mu x} (1 - \mathbb{P}[W_{n+1} = 0 \mid W_n > 0]),\end{aligned}\tag{3.8}$$

where we applied (3.7) with $E = \{W_n > 0\}$ in the final step. We obtain the probability that appears at the right-hand side of (3.8) as follows.

$$\begin{aligned}\mathbb{P}[W_{n+1} = 0 \mid W_n > 0] &= \mathbb{P}[B_{n+1} \leq A_n + W_n \mid W_n > 0] \\ &= \int_0^\infty \mathbb{P}[B_{n+1} \leq A_n + x \mid W_n > 0] \mu e^{-\mu x} dx,\end{aligned}$$

where we applied (3.6) to W_n in the last step. Since B_{n+1} is independent of W_n , we have that the previous equation becomes for $n \geq 2$,

$$\begin{aligned}\mathbb{P}[W_{n+1} = 0 \mid W_n > 0] &= \int_0^\infty \int_0^\infty (1 - e^{-\mu y} e^{-\mu x}) \mu e^{-\mu x} dx dF_A(y) \\ &= 1 - \int_0^\infty e^{-\mu y} \int_0^\infty \mu e^{-2\mu x} dx dF_A(y) = 1 - \frac{1}{2} \alpha(\mu).\end{aligned}\tag{3.9}$$

Combining (3.8) and (3.9) we have that, for $n \geq 2$, the third term at the right-hand side of (3.2) is given by

$$\mathbb{P}[W_{n+1} > x \mid W_n > 0] = \frac{1}{2} \alpha(\mu) e^{-\mu x}.\tag{3.10}$$

The last term we need to compute in order to obtain the transient distribution of the waiting times is the probability that the n -th waiting time is equal to zero, cf. (3.2). From (3.1) we readily have that $\mathbb{P}[W_2 = 0] = 1 - e^{-\mu w_1} \alpha(\mu)$. Moreover, for $n \geq 2$, we have that

$$\mathbb{P}[W_{n+1} = 0] = \mathbb{P}[W_{n+1} = 0 \mid W_n = 0] \mathbb{P}[W_n = 0] + \mathbb{P}[W_{n+1} = 0 \mid W_n > 0] \mathbb{P}[W_n > 0],$$

which implies that (cf. (3.3) and (3.9))

$$\begin{aligned}\mathbb{P}[W_{n+1} = 0] &= (1 - \alpha(\mu)) \mathbb{P}[W_n = 0] + \left(1 - \frac{1}{2} \alpha(\mu)\right) (1 - \mathbb{P}[W_n = 0]) \\ &= 1 - \frac{1}{2} \alpha(\mu) - \frac{1}{2} \alpha(\mu) \mathbb{P}[W_n = 0].\end{aligned}$$

This gives a first order recursion for $\mathbb{P}[W_{n+1} = 0]$. With simple manipulations it is easy to show that the solution to this recursion for $n \geq 2$ is given by,

$$\mathbb{P}[W_{n+1} = 0] = \frac{2 - \alpha(\mu)}{2 + \alpha(\mu)} + \left(-\frac{\alpha(\mu)}{2}\right)^{n-1} \left(\mathbb{P}[W_2 = 0] - \frac{2 - \alpha(\mu)}{2 + \alpha(\mu)}\right), \quad n \geq 2.\tag{3.11}$$

Summing up the results we obtained in Equations (3.2), (3.4), (3.10), (3.11), we have that the distribution of W_{n+1} is given by the following theorem.

Theorem 3. *If $W_1 = w_1$, then for every $n \geq 1$ the time-dependent distribution of the waiting times is given by*

$$\mathbb{P}[W_{n+1} \leq x] = 1 - e^{-\mu x} \left[\frac{2\alpha(\mu)}{2 + \alpha(\mu)} + \left(-\frac{\alpha(\mu)}{2} \right)^{n-1} \left(\frac{2 - \alpha(\mu)}{2 + \alpha(\mu)} - \mathbb{P}[W_2 = 0] \right) \right], \quad (3.12)$$

with $\mathbb{P}[W_2 = 0] = 1 - e^{-\mu w_1} \alpha(\mu)$.

Note that (3.12) is also valid for $n = 1$, since in this case (3.12) simplifies to $\mathbb{P}[W_2 \leq x] = 1 - \mathbb{P}[W_2 > 0]e^{-\mu x}$, which is consistent with (3.1). Naturally, the term in the brackets at the right-hand side of (3.12) is the probability $\mathbb{P}[W_{n+1} > 0]$.

In the alternating service example given in the introduction, it is reasonable to assume that the server has to wait for a full preparation time at the beginning, implying that $W_1 = B_1$. In this case, it is easy to show that $\mathbb{P}[W_2 = 0] = \mathbb{P}[B_2 \leq A_1 + B_1] = 1 - \frac{1}{2}\alpha(\mu)$, which yields the following corollary.

Corollary 1. *If $W_1 = B_1$, then for every $n \geq 1$ the time-dependent distribution of the waiting times is given by*

$$\mathbb{P}[W_{n+1} \leq x] = 1 - e^{-\mu x} \left[\frac{2\alpha(\mu)}{2 + \alpha(\mu)} + \left(-\frac{\alpha(\mu)}{2} \right)^{n-1} \left(\frac{1}{2}\alpha(\mu) - \frac{2\alpha(\mu)}{2 + \alpha(\mu)} \right) \right],$$

Another result we can infer from Theorem 3 is the speed of convergence of the time-dependent distribution $\mathbb{P}[W_n \leq x]$ towards the steady-state distribution $\mathbb{P}[W \leq x]$. It is clear from (3.12) that this speed of convergence is geometrically fast at rate $\frac{1}{2}\alpha(\mu)$. It is interesting to observe that this rate is twice as fast as predicted by the upper bound (obtained by a coupling argument) in [14]. In Lindley's recursion this speed of convergence towards steady state is closely related to the tail behaviour of the cycle length, see for example [2, Chapter XIII]. In the next section, we see that the same constant $\frac{1}{2}\alpha(\mu)$ appears in a crucial way in the distribution of C .

3.2 The distribution of the cycle length

As we have mentioned before, $\{W_n\}$ is a regenerative process; regeneration occurs at times when $W_n = 0$. Let C be the random variable describing the length of a generic regeneration cycle, i.e.,

$$C = \inf\{k : W_{1+k} = 0 \mid W_1 = 0\}.$$

The main goal of this section is to derive the distribution of C . By definition, we have that

$$\mathbb{P}[C = n] = \mathbb{P}[W_{n+1} = 0, W_n > 0, \dots, W_2 > 0 \mid W_1 = 0].$$

The main result of this section is the following theorem.

Theorem 4. *Let C be the length of a regeneration cycle. Then the distribution of C is given by*

$$\mathbb{P}[C = n] = \begin{cases} 1 - \alpha(\mu) & n = 1 \\ \left[1 - \frac{1}{2}\alpha(\mu)\right] \left[\frac{1}{2}\alpha(\mu)\right]^{n-2} \alpha(\mu) & n \geq 2. \end{cases} \quad (3.13)$$

Proof. For $n = 1$, we readily have from (3.3) that

$$\mathbb{P}[C = 1] = \mathbb{P}[W_2 = 0 \mid W_1 = 0] = 1 - \alpha(\mu). \quad (3.14)$$

For $n \geq 2$ we shall prove our assertion by induction. For $n = 2$ we have that

$$\begin{aligned} \mathbb{P}[C = 2] &= \mathbb{P}[W_3 = 0, W_2 > 0 \mid W_1 = 0] \\ &= \mathbb{P}[B_3 \leq A_2 + W_2 \mid W_2 > 0, W_1 = 0] \mathbb{P}[W_2 > 0 \mid W_1 = 0]. \end{aligned}$$

Furthermore,

$$\mathbb{P}[B_3 \leq A_2 + W_2 \mid W_2 > 0, W_1 = 0] = \int_0^\infty \mathbb{P}[B_3 \leq A_2 + x \mid W_2 > 0, W_1 = 0] \mu e^{-\mu x} dx,$$

since (3.5) implies that $\mathbb{P}[W_2 \leq x \mid W_2 > 0, W_1 = 0] = 1 - e^{-\mu x}$. Moreover, since B_3 and A_2 are independent of W_2 and W_1 , we have now that

$$\mathbb{P}[C = 2] = \mathbb{P}[W_2 > 0 \mid W_1 = 0] \int_0^\infty \int_0^\infty (1 - e^{-\mu y} e^{-\mu x}) \mu e^{-\mu x} dx dF_A(y) = \alpha(\mu) \left[1 - \frac{1}{2} \alpha(\mu) \right],$$

which satisfies (3.13) for $n = 2$.

Next, assume that

$$\mathbb{P}[C = k] = \left[1 - \frac{1}{2} \alpha(\mu) \right] \left[\frac{1}{2} \alpha(\mu) \right]^{k-2} \alpha(\mu) \quad (3.15)$$

for all $k \leq n$, with $n \geq 2$. To complete the proof, we must show that

$$\mathbb{P}[C = n + 1] = \left[1 - \frac{1}{2} \alpha(\mu) \right] \left[\frac{1}{2} \alpha(\mu) \right]^{n-1} \alpha(\mu).$$

Since $\mathbb{P}[C \geq n + 1] = 1 - \mathbb{P}[C \leq n]$, (3.14) and (3.15) imply that

$$\mathbb{P}[C \geq n + 1] = \left[\frac{1}{2} \alpha(\mu) \right]^{n-1} \alpha(\mu).$$

Therefore, we have that

$$\begin{aligned} \mathbb{P}[C = n + 1] &= \mathbb{P}[W_{n+2} = 0, W_{n+1} > 0, \dots, W_2 > 0 \mid W_1 = 0] \\ &= \mathbb{P}[W_{n+2} = 0 \mid W_{n+1} > 0, \dots, W_2 > 0, W_1 = 0] \mathbb{P}[W_{n+1} > 0, \dots, W_2 > 0 \mid W_1 = 0] \\ &= \mathbb{P}[W_{n+2} = 0 \mid W_{n+1} > 0, \dots, W_2 > 0, W_1 = 0] \mathbb{P}[C \geq n + 1] \\ &= \mathbb{P}[W_{n+2} = 0 \mid W_{n+1} > 0, \dots, W_2 > 0, W_1 = 0] \left[\frac{1}{2} \alpha(\mu) \right]^{n-1} \alpha(\mu). \end{aligned}$$

It suffices to show that

$$\mathbb{P}[W_{n+2} = 0 \mid W_{n+1} > 0, \dots, W_2 > 0, W_1 = 0] = 1 - \frac{1}{2} \alpha(\mu). \quad (3.16)$$

It tempting to think that this is a trivial application of the Markov property, but this is of course not the case. We have that

$$\begin{aligned} & \mathbb{P}[W_{n+2} = 0 \mid W_{n+1} > 0, \dots, W_2 > 0, W_1 = 0] \\ &= \mathbb{P}[B_{n+2} \leq A_{n+1} + W_{n+1} \mid W_{n+1} > 0, \dots, W_2 > 0, W_1 = 0] \\ &= \int_0^\infty \mathbb{P}[B_{n+2} \leq A_{n+1} + x \mid W_{n+1} > 0, \dots, W_2 > 0, W_1 = 0] \mu e^{-\mu x} dx, \end{aligned} \quad (3.17)$$

since

$$\mathbb{P}[W_{n+1} \leq x \mid W_{n+1} > 0, \dots, W_2 > 0, W_1 = 0] = 1 - e^{-\mu x},$$

which follows from (3.7) with $E = \{W_n > 0, \dots, W_2 > 0\}$ and $w_1 = 0$. Equation (3.17) yields

$$\begin{aligned} & \mathbb{P}[W_{n+2} = 0 \mid W_{n+1} > 0, \dots, W_2 > 0, W_1 = 0] \\ &= \int_0^\infty \int_0^\infty (1 - e^{-\mu y} e^{-\mu x}) \mu e^{-\mu x} dx dF_A(y) = 1 - \frac{1}{2} \alpha(\mu), \end{aligned}$$

which is exactly Equation (3.16) that remained to be proven. \square

3.3 The covariance function

We are interested in the covariance between two waiting times. By definition, we have that

$$\text{cov}[W_n, W_{n+k}] = \mathbb{E}[W_n W_{n+k}] - \mathbb{E}[W_n] \mathbb{E}[W_{n+k}].$$

The terms $\mathbb{E}[W_n]$ and $\mathbb{E}[W_{n+k}]$ can be directly computed, for example, from Theorem 3. For the expectation of the product of the two waiting times we have that

$$\mathbb{E}[W_n W_{n+k}] = \int_0^\infty w \mathbb{E}[W_{n+k} \mid W_n = w] d\mathbb{P}[W_n \leq w].$$

Write

$$\begin{aligned} \mathbb{E}[W_{n+k} \mid W_n = w] &= \mathbb{E}[W_{n+k} \mid W_{n+k} > 0, W_n = w] \mathbb{P}[W_{n+k} > 0 \mid W_n = w] \\ &= \frac{1}{\mu} \mathbb{P}[W_{1+k} > 0 \mid W_1 = w], \end{aligned}$$

where in the last step, we applied the Markov property as well as the fact that W_{n+k} , given that $W_n = w$ and $W_{n+k} > 0$, is exponentially distributed with rate μ . The latter follows from (3.7). Thus, in order to compute the covariance between W_n and W_{n+k} , we need the distribution of W_{1+k} , conditioned on $W_1 = w$. This distribution has been derived in Theorem 3, from which it follows that

$$\mathbb{P}[W_{1+k} > 0 \mid W_1 = w] = \frac{2\alpha(\mu)}{2 + \alpha(\mu)} \left[1 - \left(-\frac{\alpha(\mu)}{2} \right)^{k-1} \right] + \left(-\frac{\alpha(\mu)}{2} \right)^{k-1} \alpha(\mu) e^{-\mu w}.$$

Combining the last three equations, we obtain

$$\begin{aligned}
\mathbb{E}[W_n W_{n+k}] &= \frac{1}{\mu} \int_0^\infty w \frac{2\alpha(\mu)}{2 + \alpha(\mu)} \left[1 - \left(-\frac{\alpha(\mu)}{2} \right)^{k-1} \right] d\mathbb{P}[W_n \leq w] \\
&\quad + \frac{1}{\mu} \left(-\frac{\alpha(\mu)}{2} \right)^{k-1} \alpha(\mu) \int_0^\infty w e^{-\mu w} d\mathbb{P}[W_n \leq w] \\
&= \frac{\mathbb{E}[W_n]}{\mu} \frac{2\alpha(\mu)}{2 + \alpha(\mu)} \left[1 - \left(-\frac{\alpha(\mu)}{2} \right)^{k-1} \right] \\
&\quad + \left(-\frac{\alpha(\mu)}{2} \right)^{k-1} \alpha(\mu) \int_0^\infty w e^{-\mu w} \mu e^{-\mu w} \frac{\mathbb{P}[W_n > 0]}{\mu} dw \\
&= \frac{\mathbb{E}[W_n]}{\mu} \frac{2\alpha(\mu)}{2 + \alpha(\mu)} \left[1 - \left(-\frac{\alpha(\mu)}{2} \right)^{k-1} \right] - \frac{\mathbb{E}[W_n]}{2\mu} \left(-\frac{\alpha(\mu)}{2} \right)^k.
\end{aligned}$$

Note that the above expression is valid only for $n \geq 2$, since we have substituted $d\mathbb{P}[W_n \leq w]$ by $\mu e^{-\mu w} \mathbb{P}[W_n > 0] dw$, cf. Theorem 3. However, if we further assume that W_1 , given that $W_1 > 0$, is (like all other W_n 's) exponentially distributed with rate μ then the above expression is valid for $n = 1$ too.

All that is left in order to compute the covariance of W_n and W_{n+k} is to compute $\mathbb{E}[W_n]$ and $\mathbb{E}[W_{n+k}]$. To this end, note that for $k \geq 1$

$$\begin{aligned}
\mathbb{E}[W_{n+k}] &= \mathbb{E}[W_{n+k} \mid W_{n+k} > 0] \mathbb{P}[W_{n+k} > 0] = \frac{1}{\mu} \mathbb{P}[W_{n+k} > 0] \\
&= \frac{1}{\mu} \int_0^\infty \mathbb{P}[W_{n+k} > 0 \mid W_n = w] d\mathbb{P}[W_n \leq w] \\
&= \frac{1}{\mu} \frac{2\alpha(\mu)}{2 + \alpha(\mu)} \left[1 - \left(-\frac{\alpha(\mu)}{2} \right)^{k-1} \right] + \frac{\alpha(\mu)}{\mu} \left(-\frac{\alpha(\mu)}{2} \right)^{k-1} \int_0^\infty e^{-\mu w} d\mathbb{P}[W_n \leq w] \\
&= \frac{1}{\mu} \frac{2\alpha(\mu)}{2 + \alpha(\mu)} \left[1 - \left(-\frac{\alpha(\mu)}{2} \right)^{k-1} \right] + \frac{\alpha(\mu)}{\mu} \left(-\frac{\alpha(\mu)}{2} \right)^{k-1} \left(\mathbb{P}[W_n = 0] + \frac{\mathbb{P}[W_n > 0]}{2} \right)
\end{aligned}$$

which implies that

$$\begin{aligned}
\mathbb{E}[W_n] \mathbb{E}[W_{n+k}] &= \frac{\mathbb{E}[W_n]}{\mu} \frac{2\alpha(\mu)}{2 + \alpha(\mu)} \left[1 - \left(-\frac{\alpha(\mu)}{2} \right)^{k-1} \right] \\
&\quad - \frac{\mathbb{E}[W_n]}{\mu} \left(-\frac{\alpha(\mu)}{2} \right)^k (2\mathbb{P}[W_n = 0] + \mathbb{P}[W_n > 0]).
\end{aligned}$$

Putting everything together we obtain

$$\begin{aligned}
\text{cov}[W_n, W_{n+k}] &= \mathbb{E}[W_n W_{n+k}] - \mathbb{E}[W_n] \mathbb{E}[W_{n+k}] \\
&= \frac{\mathbb{E}[W_n]}{\mu} (2\mathbb{P}[W_n = 0] + \mathbb{P}[W_n > 0]) \left(-\frac{\alpha(\mu)}{2} \right)^k - \frac{\mathbb{E}[W_n]}{2\mu} \left(-\frac{\alpha(\mu)}{2} \right)^k. \quad (3.18)
\end{aligned}$$

Simplifying this formula, using $\mathbb{P}[W_n = 0] + \mathbb{P}[W_n > 0] = 1$, we obtain the following theorem, which is the main result of this section.

Theorem 5. For $n \geq 2$, the covariance function between W_n and W_{n+k} , $k \geq 1$, is given by

$$\text{cov}[W_n, W_{n+k}] = \frac{\mathbb{E}[W_n]}{\mu} \left(\mathbb{P}[W_n = 0] + \frac{1}{2} \right) \left(-\frac{\alpha(\mu)}{2} \right)^k. \quad (3.19)$$

Furthermore, if W_1 , given that $W_1 > 0$, has an exponential distribution with rate μ , then the above expression is valid for $n = 1$ too.

Equation (3.19) is not valid for $k = 0$. The proof fails, for example, when computing $\mathbb{E}[W_{n+k}]$. For $k = 0$ we proceed as follows. Since W_n conditioned on $W_n > 0$ has an exponential distribution with rate μ , we have that

$$\text{var}[W_n] = \frac{\mathbb{P}[W_n > 0](2 - \mathbb{P}[W_n > 0])}{\mu^2}.$$

Theorem 5 can be applied to compute the covariance between waiting times in the alternating service example, where $W_1 = B_1$. A particularly tractable case arises when we additionally assume that $\mathbb{P}[W_1 > 0] = 2\alpha(\mu)/(2 + \alpha(\mu))$, which makes $\{W_n\}$ a stationary process. In this case, we obtain the following expression for the covariance function.

Corollary 2. If $\{W_n\}$ is stationary, then for $k \geq 1$, we have that

$$\text{cov}[W_1, W_{1+k}] = \frac{2\alpha(\mu)}{2 + \alpha(\mu)} \left(\frac{3}{2} - \frac{2\alpha(\mu)}{2 + \alpha(\mu)} \right) \frac{1}{\mu^2} \left(-\frac{\alpha(\mu)}{2} \right)^k.$$

In the next section, we compare our results with existing results for Lindley's equation.

4 A comparison to Lindley's recursion

In Section 3 we have seen that the distribution of W_n and other characteristics of $\{W_n\}$ are quite explicit if all B have an exponential distribution. Since our recursion is, up to a sign, identical to Lindley's recursion, it is interesting to compare these two recursions. Let W_n^L , $n \geq 1$, be driven by Lindley's recursion, i.e.,

$$W_{n+1}^L = \max\{0, X_{n+1} + W_n^L\},$$

with $X_{n+1} = B_{n+1} - A_n$ as defined before. Naturally, W_{n+1}^L can be interpreted as the waiting time of the $(n+1)$ -st customer in a G/G/1 queue.

In this section we compare the results we have derived so far to the analogous cases for Lindley's recursion. In other words, we assume that B_i is exponentially distributed with rate μ . For the G/M/1 queue we derive the time-dependent distribution of the waiting times in Section 4.1, and we review results on the length of the busy cycle in Section 4.2. Furthermore, we review some known results on the covariance function which are valid for the G/G/1 queue in Section 4.3.

4.1 The time-dependent distribution

The literature on time-dependent properties of Lindley's recursion and other queueing systems usually involves general expressions for the double transform $\sum_{n=0}^{\infty} r^n \mathbb{E}[e^{-sW_{n+1}^L}]$, which are derived using Spitzer's identity and the Wiener-Hopf method; see e.g. Asmussen [2] and Cohen [7].

For the distribution of W_n^L , note that the following representation holds. Let Q_n be the number of customers in the system when the n -th customer arrives, and let $Q_1 = q_1 \geq 0$. Then, since all service times (in particular the residual service times) are exponential with rate μ , W_n^L has a mixed-Erlang distribution, with mixture probabilities $\mathbb{P}[Q_n = k]$, $k = 0, \dots, n + q_1$. This result is stated as Equation (3.97) in [7, p. 229]. The probabilities $\mathbb{P}[Q_n = k]$ can be computed explicitly for the M/M/1 queue if $q_1 = 0$; see, for example, Equation (2.26) in [7, p. 185]. For the G/M/1 case, it is possible to give an expression for the generating function $\sum_{n=0}^{\infty} r^n \mathbb{P}[Q_{1+n} = j]$ if $q_1 = 0$, see the equation below (3.72) in [7, p. 221]. We conclude that the mixed-Erlang representation for the distribution of W_n^L is not very explicit.

We now give an alternative form of the distribution of W_n^L , which we could not find in the literature and which we derive by means of some simple probabilistic arguments. First, let G be an integer-valued random variable independent of everything else with $\mathbb{P}[G = n] = (1 - r)r^n$, $n \geq 0$. By conditioning on G , we have for $r \in (0, 1)$ that $\mathbb{P}[W_{G+1}^L > x] = (1 - r)f^L(r, x)$, with

$$f^L(r, x) = \sum_{n=0}^{\infty} r^n \mathbb{P}[W_{n+1}^L > x]$$

the generating function of $\mathbb{P}[W_{n+1}^L > x]$. Thus, to get an expression for $f^L(r, x)$, it suffices to obtain the distribution of W_{G+1}^L .

For this, we use two more probabilistic ideas. Assume that $W_1^L = 0$. Then we have that $W_{n+1}^L \stackrel{\mathcal{D}}{=} \max_{k=0, \dots, n} S_k$, with $S_0 = 0$ and $S_n = X_1 + \dots + X_n$; see e.g. Asmussen [2, Chapter 8]. Finally, we reduce the problem to computing the distribution of the all-time maximum of a related random walk. For this, we define an i.i.d. sequence of random variables A'_i , $i \geq 1$ as follows. For any i we let $A'_i = A_i$ with probability r and $A'_i = \infty$ with probability $1 - r$. We see that we can interpret G as the first value of i such that $A'_i = \infty$. Define $S'_n = X'_1 + \dots + X'_n$, with $X'_i = B_{i+1} - A'_i$. Since $S'_n = -\infty$ if $n \geq G$ and $S_n = S'_n$ if $n < G$, it follows that

$$W_{G+1}^L \stackrel{\mathcal{D}}{=} \max_{k=0, \dots, G} S_k \stackrel{\mathcal{D}}{=} \max_{k \geq 1} S'_k =: M_r.$$

We see that W_{G+1}^L has the same distribution as the supremum M_r of a random walk S'_n , $n \geq 1$ with exponentially distributed upward jumps. For such random walks it is straightforward to derive an analogue of Theorem 5.8 in [2, p. 238] (adapted to the case where $X'_k = -\infty$ with positive probability) which yields that

$$\mathbb{P}[M_r > x] = (1 - \eta(r)/\mu) e^{-\eta(r)x},$$

with $\eta(r)$ the unique positive solution of the equation

$$\frac{\mu}{\mu - \eta(r)} \mathbb{E}[e^{-\eta(r)A'_1}] = 1. \quad (4.1)$$

We conclude that, for the G/M/1 queue,

$$(1 - r) \sum_{n=0}^{\infty} r^n \mathbb{P}[W_{n+1}^L > x] = (1 - \eta(r)/\mu) e^{-\eta(r)x}.$$

The only case where $\eta(r)$ has an explicit expression seems to be the case where A is exponentially distributed with rate λ . In that case we get

$$\eta(r) = \frac{1}{2} \left(\mu - \lambda + \sqrt{(\lambda - \mu)^2 + 4\lambda\mu(1-r)} \right).$$

If in addition $\rho = \lambda/\mu = 1$, then $\eta(r)$ further simplifies to $\mu\sqrt{1-r}$. Using the power series expansions

$$e^y = \sum_{i=0}^{\infty} \frac{y^i}{i!}, \quad \text{and} \quad (1-r)^a = \sum_{j=0}^{\infty} \binom{a}{j} (-r)^j,$$

we then see that

$$\begin{aligned} f^L(r, x) &= \frac{1}{1-r} (1 - \sqrt{1-r}) e^{-\mu\sqrt{1-r}x} \\ &= \sum_{n=0}^{\infty} \frac{(-\mu x)^n}{n!} \sum_{k=0}^{\infty} \left(\binom{\frac{n}{2}-1}{k} - \binom{\frac{n-1}{2}}{k} \right) (-r)^k \\ &= \sum_{k=0}^{\infty} r^k (-1)^k \sum_{n=0}^{\infty} \frac{(-\mu x)^n}{n!} \left(\binom{\frac{n}{2}-1}{k} - \binom{\frac{n-1}{2}}{k} \right). \end{aligned}$$

By identifying this expression as a power series in k we see that, for the M/M/1 queue with $\rho = 1$,

$$\mathbb{P}[W_{k+1}^L > x] = (-1)^k \sum_{n=0}^{\infty} \frac{(-\mu x)^n}{n!} \left(\binom{\frac{n}{2}-1}{k} - \binom{\frac{n-1}{2}}{k} \right),$$

with $\binom{a}{j} = (\prod_{k=0}^{j-1} (a-k))/j!$. We compare this expression to the distribution of W_{n+1} in the M/M case with $W_1 = 0$ and $\lambda = \mu$. Using Theorem 3, and $\alpha(\mu) = \lambda/(\lambda + \mu) = 1/2$, and $\mathbb{P}[W_2 = 0 \mid W_1 = 0] = 1/2$, we get

$$\mathbb{P}[W_{n+1} > x \mid W_1 = 0] = \left[\frac{2}{5} - \frac{1}{10} \left(-\frac{1}{4} \right)^n \right] e^{-\mu x}.$$

In our opinion, the distribution of W_n^L is less explicit than the distribution of W_n . We have shown in Theorem 3 that the distribution of W_n is a simple mixture of an atom at zero and an exponential distribution, while it seems that the distribution of W_n^L can only be represented by its generating function, or by a mixed-Erlang representation of which the mixture probabilities are given by a generating function.

4.2 The busy cycle

For the busy cycle $C^L = \inf\{n \geq 1 : W_{n+1}^L = 0 \mid W_1^L = 0\}$, the generating function can be extracted from Equation (3.89) in Cohen [7, p. 226]. Translated to our notation, we have

$$\mathbb{E}[r^{C^L}] = \frac{r - \lambda(r)}{1 - \lambda(r)},$$

with $\lambda(r)$ the root with the smallest absolute value in the unit circle of the function $z - r\alpha(\mu(1-z))$. It can be shown that $\lambda(r) = 1 - \eta(r)/\mu$, with $\eta(r)$ determined by (4.1). This implies that

$$\mathbb{E}[r^{C^L}] = \frac{\eta(r) - \mu(1-r)}{\eta(r)}.$$

For the M/M/1 queue with load ρ , an explicit expression is available, see for example Equation (2.43) in Cohen [7, p. 190], which states that

$$\mathbb{P}[C^L = n] = \frac{1}{2n-1} \binom{2n-1}{n} \frac{\rho^{n-1}}{(1+\rho)^{2n-1}}.$$

To the best of our knowledge, there is no explicit expression available for the distribution of C^L for the G/M/1 queue. Thus, the difference in tractability between Lindley's recursion and our recursion is clear, cf. Theorem 4.

4.3 The covariance function

The literature on the covariance function of the waiting times for the single-server queue seems to be sporadic. For the G/G/1 queue, Daley [8] and Blomqvist [5, 6] give some general properties. In particular, in [8] it is shown that the serial correlation coefficients of a stationary sequence of waiting times are non-negative and decrease monotonically to zero. Comparing these results to the ones we have obtained in Section 2.2, we first observe that the qualitatively different result of Theorem 1 (showing non-monotonicity) is not surprising. It is rather a natural effect of the minus sign that appears in front of W_n in (1.1). However, the condition for Theorem 2 to hold is that $\mathbb{E}[B] < \infty$, which is less restrictive than demanding that the third moment of the service times is finite, as is the case for Lindley's recursion, cf. [8]. Furthermore, from Theorem 2 we immediately have that the infinite sum of all correlations is finite. For Lindley's recursion, the finiteness of the third moment of B is not sufficient to guarantee this. Even in this case, the series may be converging so slowly to zero, that the sum is infinite. As it is stated in Theorem 2 of [8], what is necessary and sufficient is that the fourth moment of B is also finite.

For the GI/M/1 queue Pakes [12] studies the covariance function of the waiting times. Theorem 1 of [12] gives the generating function of the correlation coefficients of the waiting times in the stationary GI/M/1 queue in terms of the unique positive solution of a specific functional equation. Furthermore, the correlation coefficients themselves are also given, but now in terms of the probabilities that no other waiting time than W_0 is equal to zero, up to time n . These expressions involve the probability generating function of the distribution of the number of customers served in a busy period, and are not very practical for numerical computations. Blanc [3] is concerned with the numerical inversion of the generating functions of the autocorrelations of the waiting times, as they are given in [12] and in Blomqvist [4], who derives for the M/G/1 queue results analogous to those in [12].

To summarise, the time-dependent analysis of (1.1) when A is generally distributed and B is exponential is far more easy, and leads to far more explicit results, than the analysis and the results obtained for the G/M/1 queue. In the following section we shall extend the results presented in Section 3 to Erlang, and eventually mixed-Erlang preparation times B_n .

5 Exact solution for Erlang preparation times

In this section we assume that, for all n , the i.i.d. random variables B_n follow an Erlang distribution with N phases and parameter μ . Although the analysis is not as straightforward as in Section 3, the idea we shall utilise in the following is very simple. Namely, if all B_n follow an Erlang distribution, then we can completely describe the system in terms of a finite-state Markov chain. Thus, it

suffices to compute the one-step transition probabilities of this Markov chain. This is done in Section 5.1, and is applied to show that W_n has a mixed-Erlang distribution. Subsequently, we derive expressions for the distribution of the cycle length and the covariance.

5.1 The time-dependent distribution

Let A be a generic service time and E_i be a random variable that follows an Erlang distribution with i phases and parameter μ , which we denote by G_i . Define F_i to be the number of remaining preparation phases that the server sees after his $(i-1)$ -th service completion, that is, at the moment he initiates his i -th waiting time. Observe that $\{F_n\}$ is a Markov chain, and let

$$p_{ij} = \mathbb{P}[F_{n+1} = j \mid F_n = i].$$

Then, for $i, j \in \{1, \dots, N\}$ we have that

$$\begin{aligned} p_{ij} &= \mathbb{P}[\text{exactly } N-j \text{ exponential phases expired during } [0, A + E_i]] \\ &= \int_0^\infty \frac{(\mu x)^{N-j}}{(N-j)!} e^{-\mu x} d\mathbb{P}[A + E_i \leq x] \\ &= \frac{(-\mu)^{N-j}}{(N-j)!} \mathcal{L}_{A+E_i}^{(N-j)}(\mu), \end{aligned}$$

where $\mathcal{L}_Y^{(N-j)}$ is the $N-j$ -th derivative of the Laplace-Stieltjes transform of a random variable Y ,

$$\begin{aligned} &= \frac{(-\mu)^{N-j}}{(N-j)!} \sum_{\ell=0}^{N-j} \binom{N-j}{\ell} \alpha^{(N-j-\ell)}(\mu) \left(\left(\frac{\mu}{\mu+s} \right)^i \right)^{(\ell)} \Big|_{s=\mu} \\ &= \frac{(-\mu)^{N-j}}{(N-j)!} \sum_{\ell=0}^{N-j} \binom{N-j}{\ell} \alpha^{(N-j-\ell)}(\mu) \left[\left(-\frac{1}{2\mu} \right)^\ell \frac{(i+\ell-1)!}{2^i(i-1)!} \right] \\ &= \frac{(-\mu)^{N-j}}{2^i} \sum_{\ell=0}^{N-j} \binom{i+\ell-1}{i-1} \frac{\alpha^{(N-j-\ell)}(\mu)}{(N-j-\ell)!} \left(-\frac{1}{2\mu} \right)^\ell. \end{aligned}$$

Furthermore, for $j \in \{1, \dots, N\}$ we have that

$$\begin{aligned} p_{0j} &= \mathbb{P}[\text{exactly } N-j \text{ exponential phases expired during } [0, A)] \\ &= \frac{(-\mu)^{N-j}}{(N-j)!} \alpha^{(N-j)}(\mu). \end{aligned}$$

The rest of the transition probabilities can be computed by the relations

$$p_{00} = 1 - \sum_{i=0}^{N-1} \frac{(-\mu)^i}{i!} \alpha^{(i)}(\mu) \quad \text{and} \quad p_{i0} = 1 - \sum_{j=0}^{N-1} \frac{(-\mu)^j}{2^i} \sum_{\ell=0}^j \binom{i+\ell-1}{i-1} \frac{\alpha^{(j-\ell)}(\mu)}{(j-\ell)!} \left(-\frac{1}{2\mu} \right)^\ell.$$

Let $\mathbf{P} = (p_{ij})$ be the transition matrix and define $G_0(x) = 1$. Then, the distribution of W_n is given by

$$\begin{aligned}\mathbb{P}[W_n \leq x] &= \sum_{i=0}^N \mathbb{P}[W_n \leq x \mid F_n = i] \mathbb{P}[F_n = i] \\ &= \sum_{i=0}^N G_i(x) \mathbb{P}[F_n = i].\end{aligned}$$

Let $\varpi_{n,i} = \mathbb{P}[F_n = i \mid W_1 = w]$ and ϖ_n be the column-vector $(\varpi_{n,0}, \dots, \varpi_{n,N})^T$. Then

$$\varpi_n = \mathbf{P} \varpi_{n-1} = \mathbf{P}^{n-2} \varpi_2. \quad (5.1)$$

It remains to compute ϖ_2 . In the same way we computed p_{ij} we get, for $j \geq 1$,

$$\begin{aligned}\varpi_{2,j} &= \mathbb{P}[\text{exactly } N-j \text{ exponential phases expired during } [0, A+w)] \\ &= \frac{(-\mu)^{N-j}}{(N-j)!} \mathcal{L}_{A+w}^{(N-j)}(\mu) \\ &= \frac{(-\mu)^{N-j}}{(N-j)!} \sum_{\ell=0}^{N-j} \binom{N-j}{\ell} \alpha^{(N-j-\ell)}(\mu) (e^{-sw})^{(\ell)} \Big|_{s=\mu} \\ &= (-\mu)^{N-j} e^{-\mu w} \sum_{\ell=0}^{N-j} \frac{(-w)^\ell}{\ell!} \frac{\alpha^{(N-j-\ell)}(\mu)}{(N-j-\ell)!}.\end{aligned} \quad (5.2)$$

This also characterises $\varpi_{2,0}$. Putting everything together, we obtain the main result of this section.

Theorem 6. *For every $n \geq 2$, W_n has a mixed-Erlang distribution with parameters μ and $\varpi_{n,0}, \dots, \varpi_{n,N}$, i.e.*

$$\mathbb{P}[W_n \leq x \mid W_1 = w] = \sum_{i=0}^N \varpi_{n,i} G_i(x),$$

with ϖ_n given by Equations (5.1) and (5.2).

It is interesting to note that W_n has a phase-type distribution with $N+1$ phases for all $n \geq 2$. This is strikingly different from Lindley's recursion. In Section 4.1, we saw that for the G/M/1 queue, W_n^L has a mixed-Erlang distribution with at most $n+1$ phases, which is unbounded in n .

5.2 The distribution of the cycle length

As before, define the cycle length C to be given by

$$C = \inf\{k \geq 1 : W_{1+k} = 0 \mid W_1 = 0\} = \inf\{k \geq 1 : F_{1+k} = 0 \mid F_1 = 0\}.$$

Of course, it always holds that $\mathbb{P}[C = 1] = \mathbb{P}[X \leq 0]$. For $n \geq 1$, we have that $\mathbb{P}[C = n+1] = \mathbb{P}[C > n] - \mathbb{P}[C > n+1]$ and

$$\mathbb{P}[C > n+1] = \sum_{i_0=1}^N \mathbb{P}[F_{n+1} = i_0, F_2 \cdot \dots \cdot F_n > 0 \mid F_1 = 0]. \quad (5.3)$$

Let $t_{0,i_0}^{(n)}$ be the probability that, conditioning on the fact that $F_1 = 0$, we shall go to state i_0 in n steps without passing through state 0 while doing so. Then we have that

$$t_{0,i_0}^{(n)} = \sum_{i_1=1}^N p_{i_1 i_0} t_{0,i_0}^{(n-1)} = \sum_{i_1=1}^N \sum_{i_2=1}^N \cdots \sum_{i_{n-1}=1}^N p_{i_1 i_0} p_{i_2 i_1} \cdots p_{i_{n-1} i_{n-2}} \mathbb{P}[F_2 = i_{n-1} \mid F_1 = 0]$$

Then from (5.3) we have that

$$\mathbb{P}[C > n + 1] = \sum_{i_0=1}^N \sum_{i_1=1}^N \cdots \sum_{i_{n-1}=1}^N p_{i_1 i_0} \cdots p_{i_{n-1} i_{n-2}} p_{0 i_{n-1}}. \quad (5.4)$$

So, if we define \mathbf{Q} to be the matrix that we obtain if we omit the first line and the first row of the matrix \mathbf{P} , \mathbf{q} to be the first row of \mathbf{P} apart from the first element, \mathbf{I} to be the $N \times N$ identity matrix, and \mathbf{e} to be the $N \times 1$ vector with all its entries equal to one, then (5.4) can be rewritten in a more compact form, which is done in the following theorem.

Theorem 7. *For every $n \geq 1$ we have that the distribution of the cycle length is given by*

$$\mathbb{P}[C = n + 1] = \mathbf{q} \mathbf{Q}^{n-1} (\mathbf{I} - \mathbf{Q}) \mathbf{e}.$$

5.3 The covariance function

It should be clear by now that extending the results of Section 3 to Erlang distributed preparation times is feasible, although not as straightforward as before. The results now are given implicitly, in terms of the transition matrix of a finite-state Markov chain. When computing the covariance function between W_n and W_{n+k} the calculations become more complex, and rather long and tedious. In this section we shall only outline the procedure of computing $\text{cov}[W_n, W_{n+k}]$ when, for all n , B_n follows an Erlang distribution.

As before, it suffices to calculate the expectation $\mathbb{E}[W_n W_{n+k}]$, since $\mathbb{E}[W_n]$ and $\mathbb{E}[W_{n+k}]$ can, in principle, be computed directly from the time-dependent distribution. Therefore, it suffices to compute $\mathbb{E}[W_{n+k} \mid W_n = w]$. To this end, we have that for all $k \geq 1$

$$\begin{aligned} \mathbb{E}[W_{n+k} \mid W_n = w] &= \mathbb{E}[W_{1+k} \mid W_1 = w] \\ &= \sum_{i=0}^N \mathbb{E}[W_{1+k} \mid F_{1+k} = i, W_1 = w] \mathbb{P}[F_{1+k} = i \mid W_1 = w]. \end{aligned} \quad (5.5)$$

Clearly, for any event E depending only on W_1, \dots, W_{n-1} , we have that for $n \geq 2$

$$\mathbb{P}[W_n \leq x \mid F_n = i, E] = G_i(x).$$

This equation is analogous to Equation (3.7). So (5.5) now becomes

$$\mathbb{E}[W_{n+k} \mid W_n = w] = \sum_{i=0}^N \frac{i}{\mu} \mathbb{P}[F_{1+k} = i \mid W_1 = w].$$

From Section 5.1 we have that for all $n \geq 3$, $\varpi_n = \mathbf{P}^{n-2} \varpi_2$, and the vector ϖ_2 is computed, cf. (5.2). From this, we infer that, for all $k \geq 1$, $\mathbb{P}[F_{1+k} = i \mid W_1 = w]$ is a polynomial of degree $N - i$

multiplied by $e^{-\mu w}$. Let $c_{1+k,j}$, $j = 0, \dots, N-i$ be the constants of this polynomial. Then, for $n \geq 2$ and $k \geq 1$ we have that

$$\mathbb{E}[W_n W_{n+k}] = \sum_{i=0}^N \frac{i}{\mu} \sum_{j=0}^{N-i} c_{1+k,j} \int_0^\infty w^{j+1} e^{-\mu w} d\mathbb{P}[W_n \leq w].$$

A lengthy but straightforward computation, using Theorem 6, shows that

$$\int_0^\infty w^{j+1} e^{-\mu w} d\mathbb{P}[W_n \leq w] = \left(\frac{1}{2}\right)^{j+1} \sum_{\ell=0}^N \varpi_{n,\ell} \frac{(\ell+j)!}{2^\ell (\ell-1)!}.$$

Using Theorem 6, we can also compute $\mathbb{E}[W_n]$ and $\mathbb{E}[W_{n+k}]$. Unfortunately, the resulting expression for the covariance is rather complicated and therefore omitted. We close this section with two remarks.

Remark 1. If the preparation times do not have an Erlang distribution, but a mixed-Erlang distribution, the analysis stays completely the same, except for the computation of the transition probabilities p_{ij} and $\varpi_{2,j}$ in Section 5.1.

Remark 2. The analysis in this section is also applicable to the non-alternating service model which is discussed in [15].

Acknowledgements

We are indebted to a referee for the careful reading of the manuscript and for pointing out numerous typos and unclear points. For the completion of this work we gratefully acknowledge the hospitality and support offered by EURANDOM, The Netherlands.

References

- [1] J. E. Angus. Classroom note: The inspection paradox inequality. *SIAM Review*, 39(1):95–97, March 1997.
- [2] S. Asmussen. *Applied Probability and Queues*. Springer-Verlag, New York, 2003.
- [3] J. P. C. Blanc. Numerical transform inversion for autocorrelations of waiting times. In H. Fleuren, D. den Hertog, and P. Kort, editors, *Operations Research Proceedings 2004*, pages 297–304, Tilburg, September 1–3 2004. Springer.
- [4] N. Blomqvist. The covariance function of the M/G/1 queueing system. *Skandinavisk Aktuarietidskrift*, 50:157–174, 1967.
- [5] N. Blomqvist. Estimation of waiting time parameters in the GI/G/1 queueing system, part I: General results. *Skandinavisk Aktuarietidskrift*, 51:178–197, 1968.
- [6] N. Blomqvist. Estimation of waiting time parameters in the GI/G/1 queueing system, part II: Heavy traffic approximations. *Skandinavisk Aktuarietidskrift*, 52:125–136, 1969.
- [7] J. W. Cohen. *The Single Server Queue*. North-Holland Publishing Co., Amsterdam, 1982.

- [8] D. J. Daley. The serial correlation coefficients of waiting times in a stationary single server queue. *Journal of the Australian Mathematical Society*, 8:683–699, 1968.
- [9] E. Hassini and R. G. Vickson. A two-carousel storage location problem. *Computers & Operations Research*, 30:527–539, 2003.
- [10] E. Koenigsberg. Analysis of the efficiency of carousel and tote-stacker performance. In J. White, editor, *Proceedings of the 7th International Conference on Automation in Warehousing*, pages 173–183, San Francisco, California, October 1986. Springer.
- [11] B. Noble. *Methods Based on the Wiener-Hopf Technique for the Solution of Partial Differential Equations*, volume 7 of *International Series of Monographs on Pure and Applied Mathematics*. Pergamon Press, New York, 1958.
- [12] A. G. Pakes. The serial correlation coefficients of waiting times in the stationary GI/M/1 queue. *The Annals of Mathematical Statistics*, 42(5):1727–1734, October 1971.
- [13] B. C. Park, J. Y. Park, and R. D. Foley. Carousel system performance. *Journal of Applied Probability*, 40(3):602–612, 2003.
- [14] M. Vasiou. A non-increasing Lindley-type equation. Technical Report 2005-015, Eurandom, Eindhoven, The Netherlands, 2005. Available at <http://www.eurandom.nl>.
- [15] M. Vasiou and I. J. B. F. Adan. An alternating service problem. *Probability in the Engineering and Informational Sciences*, 19(4):409–426, October 2005.
- [16] M. Vasiou and I. J. B. F. Adan. Exact solution to a Lindley-type equation on a bounded support. *Operations Research Letters*, 35(1):105–113, January 2007.
- [17] M. Vasiou, I. J. B. F. Adan, and J. Wessels. A Lindley-type equation arising from a carousel problem. *Journal of Applied Probability*, 41(4):1171–1181, December 2004.